Class Engagement Analyzer using Facial Feature Classification

Roy Manseras
Information Technology
Education Program
University of the Immaculate Conception
Davao City, Philippines
Email: roymanseras@gmail.com

Thelma Palaoag
College of Information
Technology and Computer Science
University of the Cordilleras
Baguio City, Philippines
Email: tpalaoag@gmail.com

Alvin Malicdem
College of Information Technology
Don Mariano Marcos
Memorial State University
La Union, Philippines
Email: engine0813@yahoo.com

***Abstract*—Effective education system can be evaluated through its Input-Process-Output framework implementation. Quality instruction is one of the input component indicators which includes student engagement as its binding measure. In classroom environment, facial expression are used by teachers to measure the affect state of the class. Incorporating technology in education help students prepare for life-long learning. Emerging technologies like Affective Computing are one of today's trends to improve quality instruction delivery by analyzing affect states of the students. This paper proposed a system of classifying student engagement using facial features. Conceptual framework of the study includes multiple face detection, facial action unit extraction and a classification model. Different algorithms were tested and compared to best configure the proposed predictive classification model. Varied test datasets were also used during experiments to gauge the accuracy and overall performance of this class engagement analyzer prototype.**

Keywords—Class engagement; affective computing; facial feature extraction; action units

I. Introduction

The ultimate goal of any education system is to build a nation filled with well-paid skilled professionals [11]. Evaluation of such education system can be determined through the implementation of the institution's input-process-output framework [14]. One of the important input indicators from the framework is the quality of instruction delivery and this indicator includes student engagement as an effective teaching variable [14].

Student engagement measures the level of attention, curiosity and interest of the students in the class and improved engagement is one of the top priorities of every teacher [1]. There are three incorporating components in student engagement: behavioral, cognitive, and emotional. Behavioral engagement refers to conduct and participation of the student to co-curricular and extra-curricular activities. Cognitive engagement on the other hand measures student's ability to understand key concepts and skills needed for learning. Lastly, emotional engagement deals with how student reacts towards the people that surrounds them [6]. In this paper, authors focused on emotional engagement through facial expressions [28] to determine current affect state of the students.

New approaches to education includes incorporating technologies in preparing students for life-long learning that succeed in this fast-changing environment [7]. Among the emerging disruptive technologies that could be incorporated in education is affective computing which is under the umbrella of intelligent applications/systems. In fact, the said domain is one of the Gartner Top 10 Strategic Technology Trends 2017 [23]. Affective computing is used to deal with human emotions. With the use of some tools (sensors, microphones, camera, etc.) [4], these technologies can help improve teacher's ability to adopt students' emotional states by evaluating facial expressions.

Real time detection of facial expressions has been effectively applied using machine learning algorithms with improved percentage of accuracy [2]. Some of these data-driven systems anchored their training datasets applying the concept of physiognomy, the art of reading traits [29] through facial features. Moreover, various studies were established focusing on face detection with emphasis on learning-centered states such as confusion, excitement, flow, frustrations [10], informed, inspired, persuaded, sentimental and amused [17].

Several researches have successfully implemented emotion detection through facial expressions using Facial Action Coding System (FACS) [9], [18], [21]. Ekman et. al have thoroughly examined the relationship of emotions through facial features [2] called Action Units (AUs). These facial features like forehead, nose, mouth, eyes, etc. [2] provide association of the conveyed facial expression to certain emotions and some of these AUs [22], [28], [26] were associated with student engagement.

In this paper, the researchers aim to integrate affective computing to aid teachers in gauging quality of instruction delivery by analyzing the class engagement in a classroom environment through facial feature classification.

A. Research Hypothesis

Effective classification of student engagement would only be identified by actual teachers in a classroom setting [12]. Hence, the researcher would like to prove that initial dataset labeled by teachers would be effective as basis in classification of student engagement by the Predictive Classification Model.

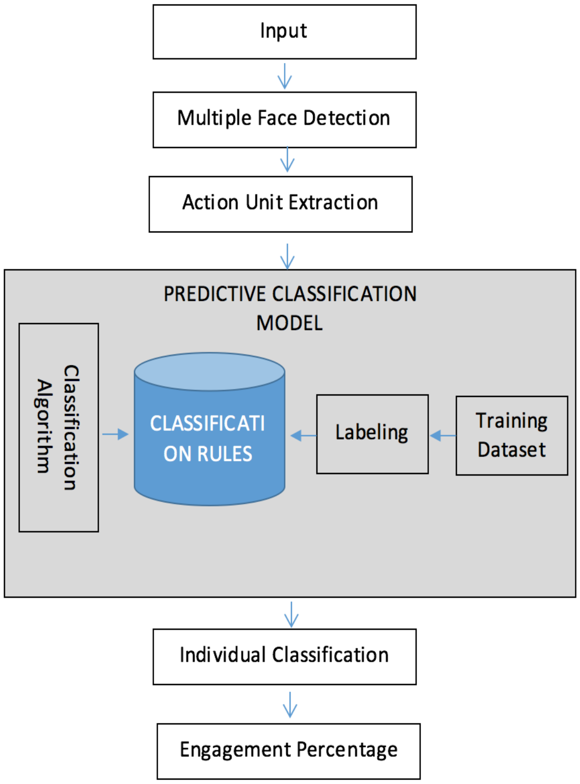

Fig. 1. Conceptual framework.

II. Methodology

A. Materials

Fig. 1 shows the conceptual framework of this study. It is composed of three major components: Input, Face Detection and Facial Feature Extraction, Predictive Classifier and Output or Feedback to user.

B. Input

The input would be a still image captured through a camera in a classroom environment where students are front facing the teacher during a class discussion. To address biased image capture of the class affect state, system setting could be set to capture an image randomly between the specified time range within class discussion. Captured image will then be loaded for the multiple face detection.

C. Multiple Face Detection

To optimize the detection of multiple faces of students in class, an advanced machine learning framework has been used for this experiment. The researchers used Face Application Program Interface which is one of the products of Microsoft Cognitive Services. The latter utilizes Microsoft Cognitive Toolkit (CNTK) [25] as its back-end. It is an open-source toolkit for deep learning algorithms [25]. This technology employs computational network (CN) which is a framework consisting of multiple components like deep neural networks (DNNs), convolutional neural networks (CNNs) [30] and others. It has been known that this toolkit evaluates deep learning algorithms faster than other tool kits [25].

TABLE I. Action Units Detected by OpenFace [3]

AU	Description
AU1	Inner brow raiser
AU2	Outer brow raiser
AU4	Brow lowerer
AU5	Upper lid raiser
AU6	Cheek raiser
AU7	Lid tighter
AU9	Nose Wrinkler
AU10	Upper lip raiser
AU12	Lip corner puller
AU14	Dimpler
AU15	Lip corner depressor
AU17	Chin raiser
AU20	Lip stretched
AU23	Lip tighter
AU25	Lips part
AU26	Jaw drop
AU28	Lip suck
AU45	Blink

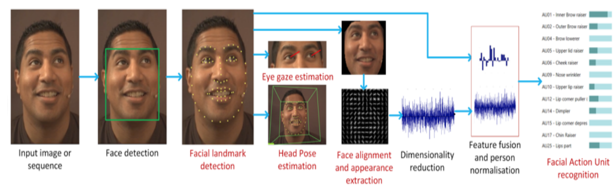

Fig. 2. Facial behavior analysis pipeline [3].

The Face Application Program Interface of Microsoft Cognitive Services offers several features like face detection, face verification, face identification, similar face searching, and face grouping [8]. Its face detection capability could detect multiple faces in a single still image.

D. Feature Extraction

After all possible faces have been detected, the system will then apply the individual faces for identifying the various facial Action Units. These AUs will then be used as parameters for the classifier.

The system will utilize OpenFace, an open source program interface that is capable of AU recognition, facial landmarks detection, head pose estimation and gaze estimation in low hardware settings [3]. Fig. 2 shows how the said program interface extracts 18 facial AUs [3]. Refer to Table 1 for the different Action Units described and its descriptions.

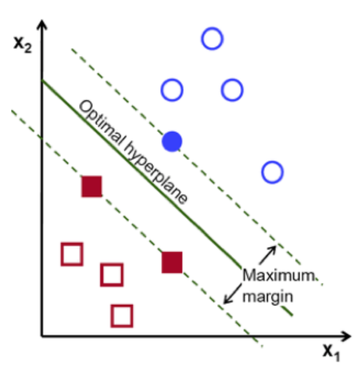

Fig. 3. Support vector machine [13].

TABLE II. 10-FOLD CROSS-VALIDATION OF THE TRAINING DATASET

Algorithm	Precision	Recall	F-Measure
Navie Bayes	0.823	0.843	0.833
Support Vector Machine	0.842	0.880	0.861
RandomForest	0.871	0.839	0.855

E. Classifier Model

To create this binary classification model, authors utilized the extracted Action Units label (true or false) and used them as parameters for classification using Support Vector Machine. In this experiment, two machine learning algorithms were tested and compared to prove that Support Vector Machine is the most appropriate for this proposed prototype.

a) Support Vector Machine (SVM): SVM has been proven to be effective in some known facial recognition systems like OpenCV [15] and CERT [5]. It is an algorithm that classifies new dataset based on the given supervised learning. SVM yields the best hyperplane that has the largest minimum distance to the given data [13]. See Fig. 3.

b) Naive Bayes: Naive Bayesian algorithm is one of the widely used classifier by various researchers because it is easy to build and applicable to large dataset. This algorithm is based on Bayes' theorem with independence assumptions between predictors [20].

c) Random Forest: It is a framework consist of various methods for classification and regression and other types. This framework works by establishing a number of decision trees at training time and display the class [27] or mean production for the individual trees [27]. Some of its best features include high accuracy even on large dataset and it could handle multiple variables without requiring reduction on attributes in high dimensional data.

The performance of the these three algorithms were estimated with cross-validation. To measure the accuracy of the trained classifier, 10-fold cross-validation (CV) was used. This process will divide the given training dataset into smaller subsets. A subset will be held as validation set to evaluate the remaining subsets of the training set in 10 folds and results are averaged over the rounds. Table 2 shows that Support Vector Machine gained well in classifying data instance as compared to the Naive Bayes and RandomForest. These results only indicate that Support Vector Machine best performs on the validation data and does not imply accuracy on unseen dataset.

F. Data Selection

Training and Validation dataset as input to the classifier model will use the Extended Cohn-Kanade (CK+) AU-Coded facial expression database [16]. This set of images consist of 123 subjects with 593 sequences [24].

In this prototype, the model will initially use labeled training dataset with the aid of human experts, in this case, two teachers as labelers. Data label for each instance could either be *Engaged* or *Not Engaged*.

Indicators used for label 'Engaged' student as discovered in [19], include expressions that denote concentration, gestures and excitement or presence of any of the following AUs as discussed in [22], [28], [26]: AU7(Lid Tighter)+AU12(Lip Corner Puller), AU5(Upper Lid Raiser), AU25(Lips Part) and AU26 (Jaw Drop); otherwise, such instance will be labeled 'Not Engaged'.

Certain procedure was applied in selecting the training dataset for the binary classification to yield better predictive model. If the assigned label differ from the two labelers (e.g. labeler1 assigned Engaged and labeler2 assigned Not Engaged), the image will be discarded. This process reduced the original dataset from 593 to 499 instances.

Final test data include three different datasets. The first two datasets were captured images from two classes with 31 and 53 students present, respectively. Last test dataset consist of combined instances from the first two datasets.

G. System Feedback

After going through the whole process of image capture to classification, the final system feedback to the user would be a percentage of labeled "Engaged" students. To get the Engagement Percentage, refer to (1).

$$\text{EP} = \frac{\text{Total True Positive}}{\text{Total Number of Instances}} \tag{1}$$

H. Accuracy Metrics

To measure the quality of the binary classifier with respect to the test dataset, Precision, Recall and F-Measures were acquired based on true positives, false positives and false negatives.

Precision refers to the number of class members classified correctly over total number of instances classified as class members. Refer to (2).

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{2}$$

Recall is comprised of the number of class members classified correctly over total number of class members. Refer to (3).

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{3}$$

F-measure is the weighted average of Precision and Recall, where score reaches its best value at 1 and worst at 0. Refer to (4).

$$\text{F-measure} = \frac{2 \times recall \times precision}{recall + precision} \tag{4}$$

III. RESULTS AND DISCUSSION

A. Face Detection and Feature Extraction

Multiple face detection using Face Application Program Interface yielded a considerable result. Out from the 31 subjects of the first class image input, only 25 where detected or 81%. And from 55 students present in the image for the second input, Face Application Program Interface detected 28 faces or equivalent of 51%. With an overall average of 66%. The researchers observed that the following factors are causes for

TABLE III. ACCURACY OF ALGORITHMS IN DATASET1 WITH 25 INSTANCES

Algortihm	Precision	Recall	F-Measure
Naive Bayes	0.941	0.800	0.865
Support Vector Machine	0.750	0.857	0.612
RandomForest	1.000	0.800	0.889

TABLE IV. ACCURACY OF ALGORITHMS IN DATASET2 WITH 28 INSTANCES

Algortihm	Precision	Recall	F-Measure
Naive Bayes	0.846	0.550	0.667
Support Vector Machine	0.833	0.750	0.789
RandomForest	0.722	0.650	0.684

TABLE V. ACCURACY OF ALGORITHMS IN DATASET3 WITH 53 INSTANCES

Algortihm	Precision	Recall	F-Measure
Naive Bayes	0.900	0.675	0.771
Support Vector Machine	0.909	0.750	0.822
RandomForest	0.853	0.725	0.784

non-detection of some instances: 1) Non-front facing position; 2) Incomplete facial features required for face detection (e.g. eyes blocked by other objects); 3) Too far from the camera causing very blurry or faded face image.

Moreover, facial Action Units were successfully extracted by OpenFace even with the very small image output cropped from the detected images of Face Application Program Interface with 64x64 image size.

B. Algorithm Performance with Test Datasets

The following explains the experiment results of the different algorithms, with emphasis on F-Measure, used with the final test datasets shown in Tables III, IV and V.

Naive Bayes algorithm lags behind the two other classifiers on all datasets except for dataset 1 where it outwits Support Vector Machine.

Moreover, RandomForest classifier started as the best classifier based on dataset 1 but remained on the second spot after increase in number of instances.

Furthermore, Support Vector Machine ranked last on the first dataset with fewer instances but as the number of instances increased, so as its accuracy performance.

IV. CONCLUSION AND FUTURE WORK

Identifying affect state, specifically student engagement, is one of the main priorities of every educator and this directed the authors to develop a tool to assess engagement percentage by analyzing facial features expressed by students during class discussions. Multiple face detection framework using Face Application Program Interface was employed to detect as many student faces as possible to gauge current engagement percentage of the whole class. The binary classifier model using Support Vector Machine was primarily set in the conceptual framework of this study. To achieve the most accuracy performance of this prototype, a comparison of Support Vector Machine to two of the most widely used binary classifiers were tested. Results show that Support Vector Machine bested RandomForest and Naive Bayesian algorithms in most of the experiments from the different test datasets.

The authors also hypothesized that the use of the human experience by the teachers would best create a model based on certain indicators to detect student engagement. This assumption was proven through the F-measure achieved by the binary classifier model using the labeled training dataset. Support Vector Machine achieved a very high accuracy rate based on F-measure on most of the test dataset experiments. These results imply the robustness of this algorithm as it achieved a very high precision rate in detecting correct classification in most of the instances.

Furthermore, other components of the framework need further studies to best improve this proposed prototype. In the case of multiple face detection, it is best to compare the Face Application Program Interface performance to other available face detection system specifically those frameworks that could address the current limitations of the CNTK. Moreover, with regards to the binary classifier model, using Support Vector Machine has proven to be effective on classifying student engagement. However, authors strongly recommend that additional faces for training dataset must be sought to improve better classification and the testing of the prototype must be set to various classroom environments to assess the efficiency of the study in different test datasets.

REFERENCES

[1] S. Abbott, 2017. [Online]. Available: http://edglossary.org/student- engagement/. [Accessed: 03- Apr- 2017].

[2] A. Al-Alwani, ”Estimation Of Mental States Using Facial Features To Improve A Students Performance In E-Learning”, in Academics World 52nd International Conference, 2016.

[3] T. Baltrusaitis, P. Robinson and L. Morency, ”OpenFace: An open source facial behavior analysis toolkit”, in 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), 2016.

[4] A. Banafa, What is Affective Computing?, OpenMind, 06-Jun-2016. [Online]. Available: https://www.bbvaopenmind.com/en/what-is-affective-computing/.

[5] G. Bradski and A. Kaehler, Learning OpenCV, 1st ed. Farnham: O'Reilly, 2013.

[6] T. Conner, ”Relationships: The Key to Student Engagement”, in International Journal of Education and Learning, 2017, pp. 13-22.

[7] ”Engaged learning for today's digital natives”, 2009. [Online]. Available: http://wwwimages.adobe.com/content/dam/Adobe/en/education/pdfs/adobe-wp-tech-education-03062009.pdf.

[8] ”Microsoft Cognitive Services - Face API”, Microsoft.com, 2017. [Online]. Available: https://www.microsoft.com/cognitive-services/en-us/face-api.

[9] J. Girard, ”Perceptions of Interpersonal Behavior are Influenced by Gender, Facial Expression Intensity, and Head Pose”, in 16th International Conference on Multimodal Interaction, 2014.

[10] J. Grafsgaard, J. Wiggins, K. Boyer, E. Wiebe and J. Lester, ”Automatically Recognizing Facial Indicators of Frustration: A Learning-Centric Analysis”, in Humaine Association Conference on Affective Computing and Intelligent Interaction, 2016.

[11] ”HTEC News”, Htecnetwork.org, 2017. [Online]. Available: http://www.htecnetwork.org/news_info.cfm?news_id=331.

[12] J. Ingosan, T. Palaog and J. Dela Cruz, ”Facial Expression Driven Mobile Learning System”, International Journal of e-Education, e-Business, e-Management and e-Learning, vol. 4, no. 1, 2014.

[13] ”Introduction to Support Vector Machines OpenCV 2.4.13.2 documentation”, Docs.opencv.org. [Online]. Available: http://docs.opencv.org/2.4/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html.

[14] Z. Kecojevic, ”Perspectives on Educational Quality: Illustrative Outcomes on Primary and Secondary Schooling in the Netherlands”, CEPS Journal: Center for Educational Policy Studies Journal, vol. 2, no. 2, p. 145, 2012.

[15] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan and M. Barlett, ”The Computer Expression Recognition Toolbox (CERT)”, in Automatic Face and Gesture Recognition and Workshops (FG 2011),IEEE, 2011, pp. 298-305.

[16] P. Lucey, J. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, ”The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression”, in Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, San Francisco, USA, 2010, pp. 94-101.

[17] D. McDuff, ”Discovering facial expressions for states of amused, persuaded, informed, sentimental and inspired”, in 18th ACM International Conference on Multimodal Interaction, ACM, 2016, pp. 71-75.

[18] Menne and B. Lugrin, ”In the Face of Emotion: A Behavioral Study on Emotions Towards a Robot Using the Facial Action Coding System”, in Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, ACM, 2017, pp. 205-206.

[19] R. Nyman, ”Indicators of student engagement: What teachers notice during introductory algebra lessons”, International Journal for Mathematics Teaching and Learning, 2015.

[20] ”Naive Bayesian”, Saedsayad.com. [Online]. Available: http://www.saedsayad.com/naive_bayesian.htm.

[21] S. Nedkov and D. Dimov, ”Emotion recognition by face dynamics”, in 14th International Conference on Computer Systems and Technologies, ACM, 2013, pp. 128-136.

[22] G. Padrn-Rivera, G. Rebolledo-Mendez and P. Parra, ”Identification of Action Units Related to Affective States in a Tutoring System for Mathematics”, Educational Technology and Society, vol. 19, no. 2, pp. 77-86, 2016.

[23] K. Panetta, ”Gartner’s Top 10 Strategic Technology Trends for 2017 - Smarter With Gartner”, Smarter With Gartner, 2017. [Online]. Available: http://www.gartner.com/smarterwithgartner/gartners-top-10-technology-trends-2017/.

[24] ”The Affect Analysis Group at Pittsburgh”, Pitt.edu. [Online]. Available: http://www.pitt.edu/ emotion/ck-spread.htm.

[25] ”The Microsoft Cognitive Toolkit - Microsoft Research”, Microsoft Research. [Online]. Available: https://www.microsoft.com/en-us/research/product/cognitive-toolkit/.

[26] A. Vail, J. Wiggins, J. Grafsgaard, K. Boyer, E. Wiebe, J. Lester, ”The Affective Impact of Tutor Questions: Predicting Frustration and Engagement”, in 9th International Conference on Educational Data Mining, 2016.

[27] M. Walker, ”Random Forests Algorithm”, Datasciencecentral.com, 2013. [Online]. Available: http://www.datasciencecentral.com/profiles/blogs/random-forests-algorithm.

[28] J. Whitehill, Z. Serpell, Y. Lin, A. Foster and J. Movellan, ”The faces of engagement: Automatic recognition of student engagementfrom facial expressions”, IEEE Transactions on Affective Computing, vol. 5, no. 1, pp. 86-98, 2014.

[29] R. Hassin and Y. Trope, ”Facing faces: Studies on the cognitive aspects of physiognomy.”, Journal of Personality and Social Psychology, Vol 78(5), May 2000, 837-852.

[30] D. Yu, A. Eversole, M. Seltzer, K. Yao, Z. Huang, B. Guenter, O. Kuchaiev, Y. Zhang, F. Seide, ... and H. Wang, ”An introduction to computational networks and the computational network toolkit”, Microsoft Technical Report MSR-TR-2014–112, 2017.